Did Protestantism promote economic prosperity via higher human capital?

Jeremy Edwards

University of Cambridge, CESifo, Munich

August 2017

Did Protestantism promote economic prosperity via higher human capital?

Jeremy Edwards
University of Cambridge and CESifo

August 2017

Abstract

This paper investigates the Becker-Woessmann (2009) argument that Protestants were more prosperous in nineteenth-century Prussia because they were more literate, a version of the Weber thesis, and shows that it cannot be sustained. The econometric analysis on which Becker and Woessman based their argument is fundamentally flawed, because their instrumental variable does not satisfy the exclusion restriction. When an appropriate instrumental-variable specification is used, the evidence from nineteenth-century Prussia rejects the human-capital version of the Weber thesis put forward by Becker and Woessmann.

JEL Codes: N33, Z12, I20, C26.
Keywords: human capital, Protestantism, economic history, instrumental variables.

Acknowledgements: I thank Tim Guinnane, Alexander Klein and Sheilagh Ogilvie for very helpful comments. I also thank Sascha Becker and Ludger Woessmann for posting their data at http://www.cesifo-group.de/ifoHome/facts/iPEHD-Ifo-Prussian-Economic-History-Database/publications.html and thus enabling me to write this paper.

1. Introduction

Did Protestantism promote economic development? Max Weber's initial formulation of such a causal link was that Protestantism promoted the work ethic and wealth accumulation, thereby encouraging the development of capitalism. Criticism of this specific argument, however, has led to the Weber thesis now being more commonly interpreted as a general claim that there is some sort of causal connection between Protestantism and economic prosperity. Becker and Woessmann (2009) (henceforth BW) provide a new interpretation of the Weber thesis in this general form. The authors argue, using data from late-nineteenth-century Prussia, that Protestantism did indeed lead to greater prosperity, the reason being that Protestants were better educated than Catholics and thus had more human capital. The human capital advantage of Protestants was an unintentional consequence of the emphasis Protestantism placed on education so that everyone could read the Bible.[1] BW do not rule out the possibility that other aspects of Protestantism, such as a work ethic, contributed to Protestants' greater prosperity, but they argue that higher human capital explains most of the difference, and possibly all.[2]

This paper shows that BW's interpretation of the Weber thesis is not supported by the Prussian data on which it is based. Their argument proceeds as follows. BW use an instrumental-variable (henceforth IV) analysis to show that there was a positive causal effect of Protestantism on literacy, their measure of human capital. They then use an IV analysis to argue that there was a positive effect of Protestantism on economic prosperity in a regression model which does not include a measure of literacy. Finally, in the absence of an IV for literacy, they assume a range of values for the possible effect of literacy on economic prosperity and use an IV analysis to estimate the effect of Protestantism on prosperity net of this assumed effect. These estimated effects vary with the assumed effect of literacy on prosperity and are not statistically significant. From this, BW conclude that most of the positive effect of Protestantism on economic prosperity is due to the greater literacy of Protestants.

[1] Becker and Woessmann (2009), 539-41.
[2] Becker and Woessmann (2009), 532.

There are two fundamental flaws in the BW argument. The first is that it is based on a misunderstanding of IV estimation. If the IV used to estimate the causal effect of Protestantism on economic prosperity in a regression model that does not include literacy satisfies the exclusion restriction required to be a valid instrument, it will give a consistent estimate of the effect of Protestantism which is not contaminated by any correlation between Protestantism and the omitted variable literacy. Indeed, if the instrument is valid, it will also give a consistent estimate of the causal effect of Protestantism in a regression model in which the dependent variable is economic prosperity net of an assumed effect of literacy on prosperity. Thus if BW's IV for Protestantism is valid, there should be no significant difference between the estimated effect of Protestantism in the regression model that takes no account of any effect of literacy on economic prosperity and the estimate in a model which assumes a particular value for the effect of literacy. Furthermore, if the instrument is valid, the estimated effect of Protestantism should not differ significantly according to the assumed value of the effect of literacy in the models in which the dependent variable is prosperity net of the assumed effect of literacy. However, BW's results do not satisfy these properties. This constitutes evidence that their IV for Protestantism is not a valid instrument in the regression models they estimate.

The second fundamental flaw in the BW argument is that if, as they show, Protestantism has a causal effect on literacy, then any variable that is correlated with Protestantism cannot be a valid IV for Protestantism in a regression model that omits literacy as a regressor. Since Protestantism is correlated with literacy because of its causal effect on literacy, any variable that is proposed as an instrument for Protestantism in such a model must necessarily be correlated with literacy, and so cannot satisfy the exclusion restriction. It is therefore unsurprising that, as pointed out above, the results BW obtain from their IV analysis provide evidence that their instrument is not valid.

In order to obtain consistent estimates of the causal effects of Protestantism and literacy on economic prosperity which allow for the causal effect of Protestantism on literacy, it is essential to have a plausible IV for literacy. Becker et al. (2011) analyse the relationship between education and industrialisation in Prussia using a dataset for the 334 Prussian counties that existed in 1849, and argue that pre-industrial

education, for which they have data, can be used as an IV for literacy. In this paper, I combine the dataset used in BW with the dataset used in Becker et al. (2011). This makes it possible to use a plausible IV for literacy and obtain consistent estimates of the causal effects of both Protestantism and literacy on economic prosperity.

The IV for Protestantism used by BW in their regression models is the distance of a county – the unit of observation in their dataset – from Wittenberg, the city where Luther's nailing of 95 theses to the door of the castle church in 1517 is conventionally taken to mark the beginning of the Protestant Reformation. In principle, distance to Wittenberg can serve as an instrument for Protestantism, but care must be taken when so using it in regression models of county economic outcomes. A central theme in the literature on nineteenth-century German industrialisation is the importance of regional effects. BW's regression models of economic outcomes in Prussia do not include any variables that measure regional effects, and hence, because distance to Wittenberg is a geographical variable, it is likely to be correlated with these omitted regional effects and thus likely to be an invalid instrument. I show that this problem is a real one: in order for distance to Wittenberg to be a valid instrument for Protestantism it is necessary for the regression model to include measures of regional effects within Prussia.

The paper therefore proceeds as follows. Section 2 of the paper outlines the BW argument. Section 3 discusses the fundamental flaws in it. Section 4 considers the importance of regional effects in Prussian economic development and shows that distance to Wittenberg is correlated with measures of these effects. Section 5 describes the data used for the main analysis in this paper. Section 6 shows that the positive causal effect of Protestantism on literacy is still present in a regression model which, in order to allay concerns about the validity of distance to Wittenberg as an instrument, includes measures of regional effects. This effect is, however, a very small one. Section 7 provides estimates of the causal effects of both Protestantism and literacy on the three measures of Prussian county economic outcomes used by BW. The results are not the same across all three measures, but they unambiguously reject BW's claim that Protestantism had a positive effect on economic prosperity which was largely due to the higher human capital of Protestants. Section 8 concludes.

2. An outline of the BW argument

BW's argument that Protestants were more prosperous than Catholics in late nineteenth-century Prussia because they had more human capital has three components. The first is the claim that, because of Luther's belief that everyone should be able to read the Bible, Protestantism promoted education, and hence Protestants had greater human capital than Catholics. To support this claim, BW use data for the 452 Prussian counties in the 1871 census to estimate regressions in which the share of literates aged 10 or over in a county's population is explained by the share of Protestants in a county's population, a number of county-level demographic variables, and a measure of the share of the county population for which literacy information is missing.[3] BW argue that the distance from a county to Wittenberg provides a way of obtaining variation in the share of Protestants in a county that is independent of economic and educational considerations.[4] Hence, they argue, the causal effect of Protestantism on literacy can be estimated by using distance to Wittenberg as an instrument for the share of Protestants in an IV regression model. BW's IV estimates show that Protestantism did have a positive causal effect on literacy, thus supporting the first component of their argument.

BW's initial arguments for the relevance of the distance to Wittenberg as an instrument for the share of Protestants were strengthened by Cantoni (2012). That paper argued that a theory of strategic neighbourhood interactions explains why distance to Wittenberg is a good predictor of the adoption of Protestantism by a territory of the Holy Roman Empire in the sixteenth and seventeenth centuries. Because of the Catholic sympathies of the Holy Roman Emperors, it was risky for territorial rulers to adopt the Reformation. If neighbouring rulers adopted the Reformation, the risk of a given ruler so doing was reduced, and hence this ruler was more likely to adopt Protestantism. Cantoni finds evidence that an increase in the proportion of neighbouring rulers who had adopted the Reformation led to an increase in the probability of a ruler subsequently adopting the Reformation. He argues that the early adoption of Protestantism by the Elector of Saxony, in which Wittenberg was

[3] County is BW's translation of the German word *Kreis*. A *Kreis* is an administrative unit which is closer to the American than to the British sense of county.
[4] Becker and Woessmann (2009), 557-63.

located, is the key factor in explaining why distance to Wittenberg predicts Protestantism.

The second component of the BW argument is the claim that Protestantism led to economic prosperity. To support this claim, BW use three different measures of economic outcomes: income tax revenue per capita in each Prussian county in 1877 (their preferred measure), the natural logarithm of average annual income of male elementary school teachers in each county in 1886, and the share of the labour force in manufacturing and service occupations in each county in 1882. IV regressions of these measures on the share of Protestants in each county (instrumented by distance to Wittenberg) and the demographic control variables show that Protestantism had a positive causal effect on all three economic outcomes. Since literacy is not included as an explanatory variable in these regressions, BW interpret the estimated coefficient of the share of Protestants as the total causal effect of Protestantism on economic outcomes, including any effect that operates because Protestantism promotes literacy.[5] These results seem to confirm the second component of the BW argument.

The third and final component of the BW argument is the claim that the positive causal effect of Protestantism on economic outcomes is largely due to the higher literacy of Protestants. To establish this claim, both literacy and Protestantism have to be included as explanatory variables in the regression analysis of economic outcomes. Literacy is an endogenous variable in the economic outcome regressions, since, according to the BW argument, it is causally related to Protestantism. Furthermore, literacy may depend on economic outcomes, and both literacy and economic outcomes are likely to depend on unobservables such as ability and work ethic. Thus an instrument for literacy is required in order to estimate its causal effect on economic outcomes. However, no instrument for literacy is available in BW's dataset. To circumvent this problem, BW assume a range of values for the causal effect of literacy which they argue to be plausible. Using these assumed values, they calculate a range of values for county economic outcomes, net of the assumed effect of literacy. IV regression of these net economic outcomes on the share of Protestants (instrumented by distance to Wittenberg) and the demographic control variables

[5] Becker and Woessmann (2009), 569.

produces estimates of the causal effect of Protestantism that control for the assumed effect of literacy. These estimates appear to show that literacy can account for most, or even all, of the positive effect of Protestantism on economic outcomes: once literacy is controlled for, the estimated effect of Protestantism is small and not statistically significant. Thus the third component of the BW argument also seems to be confirmed.

However, as the remainder of this paper will show, BW's argument is incorrect. The claim that Protestantism had positive effects on county economic outcomes that were largely due to the higher literacy of Protestants does not stand up to careful scrutiny.

3. Fundamental problems with the BW argument

The central plank in BW's argument that Protestants were more prosperous than Catholics in late-nineteenth-century Prussia because they had more human capital is laid out in section VI.C of their paper. IV regression of county economic outcomes, net of the assumed effect of literacy, on the share of Protestants (instrumented by distance to Wittenberg) and the demographic control variables produces estimates of the causal effect of Protestantism that are claimed by BW to control for the assumed effect of literacy. These estimates, given in Tables VI and VII of their paper, appear to show that literacy can account for most, or even all, of the positive effect of Protestantism on economic outcomes: once literacy is controlled for, the estimated effect of Protestantism is small and not statistically significant.

The problem with this argument is that it appears not to appreciate the standard case for IV estimation when there are omitted variables. If distance to Wittenberg is a valid IV for the share of Protestants, then a consistent estimate of the causal impact of Protestantism on economic outcomes that is uncontaminated by any correlation between Protestantism and literacy can be obtained even if literacy is omitted from the regression model. There is no need for the BW procedure of assuming plausible values for the causal effect of literacy on economic outcomes, calculating values of economic outcomes net of the assumed effect of literacy, and

estimating the effect of the share of Protestants on these net values by using distance to Wittenberg as an instrument.

To see this, suppose that the population regression model relating economic outcomes (Y_i), the share of Protestants (P_i) and the share of literates (L_i) is

$$Y_i = \alpha + \beta P_i + \gamma L_i + \varepsilon_i, i = 1,...,n \quad (1)$$

where ε_i is an error term that is uncorrelated with L_i.[6] Omitting L_i from the regression model results in

$$Y_i = \alpha + \beta P_i + e_i, i = 1,...,n \quad (2)$$

where

$$e_i = \gamma L_i + \varepsilon_i, i = 1,...,n.$$

Let Z_i be an IV which satisfies the standard relevance and exclusion restriction assumptions: $\text{cov}(Z_i, P_i) \neq 0$, $\text{cov}(Z_i.L_i) = 0$ and $\text{cov}(Z_i.\varepsilon_i) = 0$, where these are all population covariances. From equation (2)

$$\begin{aligned}\text{cov}(Z_i, Y_i) &= \text{cov}(Z_i, \alpha + \beta P_i + e_i) \\ &= \beta \,\text{cov}(Z_i, P_i) + \gamma \,\text{cov}(Z_i, L_i) + \text{cov}(Z_i, \varepsilon_i).\end{aligned}$$

Since Z_i is assumed both to be relevant and to satisfy the exclusion restriction,

$$\beta = \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, P_i)} \quad (3)$$

so that the population coefficient β is the ratio of the population covariance between Z and Y to the population covariance between Z and P.

The IV estimator of β, denoted $\hat{\beta}$, replaces the population covariances in (3) with sample covariances between Z and Y (s_{ZY}) and between Z and P (s_{ZP}):

$$\hat{\beta} = \frac{s_{ZY}}{s_{ZP}}. \quad (4)$$

The sample covariance is a consistent estimator of the population covariance and hence the IV estimator $\hat{\beta}$ is consistent:

$$\hat{\beta} \xrightarrow{p} \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, P_i)} = \beta.$$

[6] If there are additional regressors that are uncorrelated with the error term, equation (1) can be thought of as the regression equation after these additional regressors have been partialled out.

Thus, even if literacy is omitted from a regression model of county economic outcomes, the IV estimator will yield a consistent estimator of the causal effect of the share of Protestants provided that distance to Wittenberg is a valid instrument for the share of Protestants. It is only when the OLS rather than the IV estimator is used that the omission of literacy will, if there is a positive correlation between Protestantism and literacy, result in the estimated effect of Protestantism being too high.

BW use IV to estimate a regression model which relates economic outcomes net of an assumed effect of literacy to the share of Protestants. Does this approach also yield a consistent estimate of the causal effect of Protestantism on economic outcomes? Let $W_i = Y_i - \delta L_i, i = 1,...,n$ where δ is the assumed coefficient of literacy. Equation (1) can then be written as

$$W_i = \alpha + \beta P_i + (\gamma - \delta)L_i + \varepsilon_i, i = 1,...,n$$

and, since literacy is omitted from the regression model, BW estimate

$$W_i = \alpha + \beta P_i + u_i, i = 1,...,n \qquad (5)$$

where $u_i = (\gamma - \delta)L_i + \varepsilon_i$. Maintaining the assumptions that Z_i is a relevant instrument and can be excluded from (5), the same argument that led to (3) shows that

$$\beta = \frac{\text{cov}(Z_i, W_i)}{\text{cov}(Z_i, P_i)}.$$

The IV estimator of β in (5), denoted $\tilde{\beta}$, is

$$\tilde{\beta} = \frac{s_{ZW}}{s_{ZP}}$$

where s_{ZW} denotes the sample covariance between Z and W. Hence the IV estimator $\tilde{\beta}$ is consistent:

$$\tilde{\beta} \xrightarrow{p} \frac{\text{cov}(Z_i, W_i)}{\text{cov}(Z_i, P_i)} = \beta.$$

This analysis shows that if, in the regression models estimated by BW, distance to Wittenberg is a relevant instrument for Protestantism and satisfies the exclusion restriction, the following results should be expected. There should be no significant difference, at least in large samples, between the IV estimate of the effect of Protestantism on economic outcomes obtained from a model in which no

assumption is made about the effect of literacy and the IV estimate of this effect obtained from a model in which the effect of literacy is assumed to take a specific value. Furthermore, in the latter case, the IV estimate of the effect of the share of Protestants should not differ significantly whatever assumption is made about δ, the effect of literacy on economic outcomes.

BW's estimates in Tables V and VI of their paper do not exhibit these features. Equation (4) in Table V shows that the IV point estimate of the effect of the share of Protestants on income tax per capita in their regression model which omits literacy is 0.586.[7] Table VI shows that the IV point estimates of the effect of the share of Protestants in their regression models which incorporate an assumed effect of literacy on economic outcomes vary from 0.309 to -0.061. It is not correct to interpret these findings as evidence that literacy can account for most of the positive effect of Protestantism on economic outcomes. Rather, these results are evidence that distance to Wittenberg is not a valid instrument for Protestantism in BW's regression models, as I shall now show.

Suppose in the analysis above that, although the population covariance between the instrument Z and ε is zero, the population covariance between Z and L is non-zero: $\text{cov}(Z_i, L_i) \neq 0$. The instrument Z_i does not therefore satisfy the exclusion restriction. In this case

$$\frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, P_i)} = \beta + \gamma \frac{\text{cov}(Z_i, L_i)}{\text{cov}(Z_i, P_i)} \tag{6}$$

and

$$\frac{\text{cov}(Z_i, W_i)}{\text{cov}(Z_i, P_i)} = \beta + (\gamma - \delta) \frac{\text{cov}(Z_i, L_i)}{\text{cov}(Z_i, P_i)}. \tag{7}$$

Neither the IV estimator $\hat{\beta}$ nor the IV estimator $\tilde{\beta}$ are consistent in this case. The former, obtained from the regression model that omits literacy, will converge in probability to the right-hand side of (6) while the latter, obtained from the regression model in which the effect of literacy is assumed to be δ, will converge in probability to the right-hand side of (7). The two estimators should yield different results, since

[7] Income tax per capita is BW's preferred measure of county economic outcomes.

$$\hat{\beta} - \tilde{\beta} \xrightarrow{p} \delta \frac{\text{cov}(Z_i, L_i)}{\text{cov}(Z_i, P_i)}$$

and, from (7), the estimated effect of Protestantism in the regression model in which the effect of literacy is assumed to be δ will depend on the assumed value of δ.

These are exactly the features of the results in Tables V and VI of the BW paper. The BW findings are what would be expected if their IV, distance to Wittenberg, is correlated with the omitted variable literacy, and thus is not a valid IV in BW's regression models relating economic outcomes to Protestantism.

There are two reasons why distance to Wittenberg is not a valid IV for the share of Protestants in BW's regression models of county economic outcomes. One reason is not specific to distance to Wittenberg but applies to any instrument that might be used for Protestantism in models of economic outcomes that omit literacy as a regressor. The other reason applies specifically to distance to Wittenberg.

If Protestantism has a causal effect on literacy, then any variable that is sufficiently correlated with Protestantism to be a potential IV must be correlated with literacy as well. Hence it is not possible to find an instrument for the share of Protestants that will also satisfy the exclusion restriction in a regression model that omits literacy. The BW argument that Protestantism influences literacy means that any variable that is correlated with the share of Protestants must inevitably also be correlated with literacy and cannot therefore be a valid IV. The estimates obtained by BW using distance to Wittenberg as an instrument for Protestantism in regression models of economic outcomes, irrespective of whether they are or are not based on assumed values for the effect of literacy, must be inconsistent.

Even if this fundamental problem did not exist, distance to Wittenberg is unlikely to be a valid IV for the share of Protestants in BW's regression models of county economic outcomes. This is because those models do not include any regressors that reflect regional effects on economic outcomes. The importance of such effects is a central theme in the literature on German industrialisation in the nineteenth century, and any analysis of county economic outcomes must take them

into account.[8] Regional effects are very likely to be correlated with distance to Wittenberg, since both are geographical in nature. The omission of regional effects from BW's regression models mean that distance to Wittenberg is therefore very likely to be an invalid instrument, as is discussed in the next section.

4. Regional effects in the Prussian economy and their implications

Nineteenth-century Prussia consisted of territories that had been part of the Prussian state for very different lengths of time. The Duchy of Prussia was created in 1525 and was unified with Brandenburg in 1618 to become the state of Brandenburg-Prussia, which also included some small territories in the Rhineland. In 1701 this state became the Kingdom of Prussia, and during the eighteenth century it expanded by acquiring, inter alia, Pomerania, Silesia, and additional parts of Poland. In 1815 Prussia acquired all of the Rhineland, Westphalia, and various other territories, and in 1866 Prussia annexed Hannover, Hessen, and Schleswig-Holstein. Of the 452 counties in BW's database, 66 had been Prussian since 1525, 108 had become Prussian in the 1810s, and 86 had only been Prussian since 1866.

There was a great deal of variation in the institutional frameworks of different parts of Prussia and this influenced economic development throughout the nineteenth century.[9] The powers of feudal landlords remained strong in the backward eastern parts of Prussia even after the formal abolition of Prussian serfdom in 1806, and factory industrialisation here was delayed until the later nineteenth century. Despite having very dense proto-industry in the early nineteenth century, Silesian industrialisation was hampered by the desire of feudal landlords to protect their proto-industrial feudal revenues through resistance to technological improvements in linen production, in which they were supported by the Prussian state. The Rhineland was the most economically advanced part of Prussia in 1816, because the early decline in landlord power combined with extensive political fragmentation to enable proto-industries easily to cross territorial boundaries in order to locate where political and institutional conditions were least oppressive, and the institutional framework here

[8] Tipton (1976), Ogilvie (1996a).
[9] Ogilvie (1996b), 265.

remained favourable to economic development throughout the nineteenth century.[10] Acemoglu et al. (2011) construct an index of German institutional reform based on the civil code, agrarian reform, the abolition of guilds, and the abolition of serfdom. The value of this index for the Rhineland was considerably higher, and the value for Saxony was modestly higher, than for the provinces of Prussia, Brandenburg, Pomerania, Silesia and Westphalia in both 1850 and 1900.[11]

The omission of any variables measuring regional effects from BW's regression models of county economic outcomes might not result in inconsistent estimates of the effect of Protestantism if BW's instrument for Protestantism could plausibly be argued to be uncorrelated with the omitted regional effects. But BW's instrument is distance to Wittenberg, a geographical variable. As I shall now show, distance to Wittenberg is indeed correlated with measures of these regional effects, and hence it is unlikely to be a valid IV in regression models of county economic effects which omit them.

A natural way to allow for the effects of institutional frameworks that differ between Prussian provinces in regression models of county economic outcomes is to suppose that there are province fixed effects on these outcomes, which can be captured by including provincial dummy variables as regressors. The counties in BW's dataset can be categorised according to the provinces of which they were part in 1871 using the ifo Prussian Economic History Database.[12] There were 57 counties in the province of Prussia, 34 in Brandenburg, 29 in Pomerania, 27 in Posen, 63 in Silesia, 42 in Saxony, 20 in Schleswig-Holstein, 37 in Hannover, 35 in Westphalia, 35 in Hessen, 69 in the Rhineland, and 4 in Hohenzollern.

Province dummy variables on their own, however, are unlikely to capture all the effects of differing institutional frameworks on county economic outcomes, since within each province counties varied in the length of time they had been part of the Prussian state. In the Rhineland province, for example, 59 counties had become Prussian in 1815, but two had been Prussian since 1614, one since 1666, three since

[10] Ogilvie (1996a), 124-5.
[11] Acemoglu et al. (2011), Table 1, 3292.
[12] Becker et al. (2014).

1702 and one since 1713. I therefore used the year of a county's annexation by Prussia as a regressor in order to allow for possible effects of the length of time it had been Prussian on economic outcomes. In addition, I used interactions between the year of annexation and the province dummy variables to allow for the possibility that the effects of the length of time a county had been Prussian varied by province.[13]

There is also evidence that three other geographical variables - the distance of a county from Berlin, from London, and from the nearest provincial capital - influenced the industrial development of Prussian counties in the latter part of the nineteenth century (Becker et al. 2011, Edwards 2017). These variables are also likely to be correlated with distance to Wittenberg, raising further questions about the validity of BW's instrument in regression models that omit them as possible influences on county economic outcomes. BW's dataset does not contain measures of the distance of a county from London and the nearest provincial capital, but it does include the latitude and longitude of each county, from which, together with the latitude and longitude of London and the provincial capitals, I was able to calculate these two distance measures.

Table 1 reports the results of two regressions which show the association between distance to Wittenberg and the various geographical measures that are likely to affect county economic outcomes. The four observations for the province of Hohenzollern were not used in the estimation of these regressions, because consistent estimates of the fixed effect of this province cannot be obtained from such a small number of observations. The two regressions differ according to whether the demographic control variables used by BW in their analysis of the relationship between economic outcomes and Protestantism were included as regressors. The conclusions that I draw from the results in Table 1 do not depend on the presence or absence of these regressors.

For both equations in Table 1, the null hypotheses that the coefficients of the ten province dummy variables were all zero and the coefficients of the eight terms that interacted province dummies with the year of annexation were all zero were

[13] The counties in the provinces of Schleswig-Holstein and Hessen all became Prussian in 1866 so that interaction terms for these provinces were not required.

Table 1: The relationship between distance to Wittenberg and possible omitted variables in regression models of county economic outcomes, Prussia 1871

	Dependent variable: Distance to Wittenberg	
Regressors	(1.1)	(1.2)
Brandenburg	-19.895	-6.080
	(15.367)	(17.617)
Pomerania	9.695**	10.644**
	(4.387)	(4.980)
Posen	-18.142***	-18.204***
	(2.818)	(3.557)
Silesia	-59.973***	-54.168***
	(3.122)	(4.103)
Saxony	-122.011***	-107.078***
	(15.043)	(16.985)
Schleswig-Holstein	-31.516***	-27.463**
	(11.358)	(10.703)
Hannover	-79.001***	-70.497***
	(15.793)	(16.383)
Westphalia	-95.062***	-83.464***
	(18.555)	(20.492)
Hessen	-122.715***	-108.912***
	(14.421)	(15.914)
Rhineland	-100.409***	-88.142***
	(19.650)	(20.978)
Year in which annexed by Prussia	-1.027***	-0.644**
	(0.195)	(0.259)
Distance to Berlin	0.968***	0.979***
	(0.017)	(0.018)
Distance to London	0.063**	0.064**
	(0.027)	(0.027)
Distance to nearest provincial capital	-0.002	0.009
	(0.047)	(0.049)
Constant	-25.070	-222.783**
	(83.450)	(105.180)
Adjusted R^2	0.990	0.990

Notes: Number of observations for both equations is 448. Figures in parentheses are heteroscedasticity-robust standard errors. ** and *** denote significance at the 0.05 and 0.01 levels respectively. Equation (1.2) also includes the following regressors, the coefficients of which are not reported: percentage aged below 10, percentage Jewish, percentage female, percentage native residents, percentage Prussian nationality, average household size, ln (population size), population growth 1867-71, percentage blind, percentage deaf-mute, and percentage insane. See text for interpretation of estimated effects of year of annexation and province dummy variables.

strongly rejected. Thus there is clear evidence that these variables were associated with the distance of a county from Wittenberg. The results for year of annexation and province reported in Table 1 are the marginal effects for each variable. The marginal effect of year of annexation is evaluated at the mean values of the province dummies for the entire sample, while the marginal effects of the provinces are evaluated at the mean values of year of annexation for the province in question. The omitted province

dummy is that for the province of Prussia, so the province marginal effects show the difference compared to that province.

It is clear from Table 1 that distance to Wittenberg is statistically significantly associated with the variables that measure regional effects on Prussian economic development as well as the measures of the distance of a county from both Berlin and London. In many cases these associations are large. The mean value of the distance to Wittenberg for the 448 observations used in Table 1 is 324.848 kilometres. Several of the province marginal effects are more than 25 per cent of this value. The marginal effect of the year of annexation in equation (1.2) corresponds to an elasticity at sample mean values of -3.47 while that in equation (1.1) corresponds to one of -5.53. In equation (1.1) the partial correlation between the distance to Wittenberg and the distance to Berlin is 0.955 while in equation (1.2) this value is 0.954, so there is a particularly strong association between distance to Wittenberg and distance to Berlin.[14] If the regional effect and distance variables in Table 1 did have an effect on county economic outcomes, then their omission from BW's regression models of economic outcomes would mean that distance to Wittenberg was an invalid instrument for Protestantism in the BW models. The analysis which follows shows that this is indeed the case.

5. Data

In order to obtain a consistent estimate of the effect of Protestantism on economic development, it is essential that the regression model used for this purpose does not omit literacy. IV estimation of the effect of literacy is required because literacy is likely to be an endogenous variable in such a regression model, since both economic outcomes and Protestantism potentially have causal effects on literacy. BW were unable to obtain such an estimate of the effect of literacy because their dataset did not contain an instrument for it. However, Becker et al. (2011) used a dataset for the 334 Prussian counties that existed in 1849 to analyse the relationship between education and industrialisation. This dataset provides a possible solution to the

[14] The partial correlation measures the correlation between distance to Wittenberg and distance to Berlin variables once the effects of the other variables in the regression on these two variables have been partialled out.

problem of the missing instrument for literacy, because it has observations for the level of education in 1816, which Becker et al. argue can serve as an instrument for the 1871 literacy rate. In addition, this dataset allows a number of other potentially important variables to be included in the analysis of county economic outcomes.

I therefore used the ifo Prussian Economic History Database to supplement the database of 334 counties in 1849 by adding to it those variables used by BW in their analysis that were not in the Becker et al. dataset.[15] Several counties that existed in 1849 were subdivided between then and 1871, so in order to do this it was necessary to aggregate the data from 1871 to reconstruct the county structure of 1849, as recommended by Becker et al. (2014). Using a database for the 334 counties that existed in 1849 means, of course, that the analysis which follows excludes all the counties that became Prussian after 1849: thus there are no observations for the provinces of Hannover, Hessen, Hohenzollern and Schleswig-Holstein.

Table 2 presents summary statistics for the main variables from the sample of 334 counties used in the following analysis. Comparison of Table 2 with BW's Table I shows that the economic outcome measures and the literacy rate are similar in the two samples, although teacher income is rather more dispersed in the BW sample.[16] There is a somewhat greater difference in the share of Protestants between the two samples, with the mean value being 4.2 percentage points lower in the smaller sample, but this difference is not statistically significant. The main features of BW's dataset are preserved in the smaller sample.

To check that BW's analysis was replicated using only the 334 Prussian counties that existed in 1849, I estimated the IV regression models of the relationship between economic outcomes and Protestantism from their Table V on this smaller sample. The resulting estimates of the effect of the share of Protestants were similar to those of BW. The only difference worth noting was that the point estimate of the effect of Protestantism on the share of the labour force in manufacturing and services from the smaller sample, though similar in size to that of BW, was less well-

[15] Becker et al. (2014).
[16] In Table I of BW (2009), income tax per capita is measured in hundreds of marks, rather than in marks as in Table 2 of this paper. The apparent large difference is merely a matter of different units.

Table 2: Summary statistics for the main variables in the sample of 334 Prussian counties

	Mean	Standard deviation	Min	Max
Economic outcome variables				
Income tax revenue per capita	190.41	68.78	56.34	562.62
Income of male elementary school teachers	965.91	181.44	726.76	1913.21
Ln (income of male elementary school teachers)	6.86	0.17	6.59	7.56
Share of labour force in manufacturing and services	32.95	15.20	7.93	79.17
Explanatory variables of primary interest				
Share of Protestants in population 1871	59.98	38.98	0.26	99.80
Literacy rate 1871	85.10	14.18	9.51	99.33
Geographical variables				
Distance to London	940.34	327.74	416.00	1534.00
Distance to Berlin	329.27	160.59	1.00	650.04
Distance to nearest provincial capital	85.02	43.41	0.00	280.00
Distance to Wittenberg	333.20	164.74	0.00	731.46
County area (in thousand square kilometres)	0.81	0.45	0.00	2.54
Pre-industrial variables				
Paved streets 1815 (dummy variable)	0.22	0.42	0.00	1.00
Public buildings per capita 1821	0.004	0.003	0.0003	0.021
Sheep per capita 1816	0.55	0.44	0.00	2.58
Tonnage of ships per capita 1819	0.01	0.05	0.00	0.55
Looms per capita 1819	0.01	0.02	0.00	0.23
Share of farm labourers in population 1819	9.52	3.95	0.00	24.13
Share of population living in cities 1816	24.83	18.74	0.00	100.00
Year in which annexed by Prussia	1720.50	109.01	1525.00	1816.00

Notes: Number of observations for all variables except income tax per capita is 334. Income tax data for the city counties was unavailable so number of observations for this variable is 321.

determined, being significantly different from zero only at the 0.12 level. I also repeated the bounding analysis in BW's Table VI and obtained broadly similar results.[17] Restricting the analysis to the smaller sample of 334 counties for which there are observations of the share of Protestants in 1816 does not therefore alter the BW findings to any significant extent.

[17] Full details of the replication of the BW analysis on the smaller sample are available from the author on request.

6. Protestantism and literacy

The first component of the BW thesis is the argument that Protestantism had a positive causal effect on literacy. They support this claim by IV estimation of the effect of the share of Protestants on the literacy rate, using distance to Wittenberg as an instrument, in a regression which also includes several demographic variables. The results, in Table III of their paper, suggest that there is a causal effect of Protestantism on literacy, but of a modest size – the estimated elasticity of literacy with respect to Protestantism at sample mean values is 0.138.

As section 4 showed, the distance of a county to Wittenberg is correlated with measures of regional effects and the county's distance from Berlin and London. If these variables influenced the literacy rate, distance to Wittenberg would not be a valid instrument in BW's literacy regression model and the resulting estimate of the effect of Protestantism on literacy would be inconsistent. In order to investigate whether this problem arises, I estimated the following regression model:

$$L_i = \alpha_0 + \alpha_1 P_i + \alpha_2' PR_i + \alpha_3 A_i + \alpha_4' PR_i * A_i + \alpha_5' D_i + \eta_i .$$

Here *i* denotes county, L_i denotes the literacy rate, P_i denotes the share of Protestants, PR_i is a vector of dummy variables indicating the province in which a county is located, A_i denotes the year in which a county was annexed by Prussia, D_i is a vector of variables measuring the distance of the county to, respectively, Berlin, the provincial capital, and London, and η_i is an error term. I did not include the demographic control variables used by BW in their analysis. These variables – the proportions of the population in 1871 that were, respectively, aged below 10, female, Jewish, native-born, Prussian, missing educational information, blind, deaf-mute, and insane, together with the average household size, the log of population size, and population growth from 1867 to 1871 – are, in many cases, bad controls: "variables that are themselves outcome variables in the notional experiment at hand".[18]

As pointed out in section 4, the partial correlation between a county's distance to Wittenberg and its distance to Berlin is extremely high. Because of this, when

[18] Angrist and Pischke 2009, 64.

distance to Berlin was included as a regressor in the literacy regression model, distance to Wittenberg became an extremely weak instrument for the share of Protestants. To deal with this problem, I used the share of Protestants in county population in 1816 as an instrument for the share of Protestants in county population in 1871. I shall discuss the justification for so doing shortly, but for the moment I proceed on the assumption that the former is a valid instrument for the latter. I estimated the regression model for literacy using the share of Protestants in 1816 as the only instrument for the share of Protestants in 1871. I follow Becker et al. (2011) in clustering standard errors at the level of the 280 independent units of observation in 1816. This is because the data reported in the 1816 census had to be adjusted in order to construct a consistent dataset for the 334 Prussian counties as they existed in 1849, and hence the data for 1816 were not based on 334 independent observations.[19] I tested the null hypothesis that the coefficient of distance to Berlin was zero. This null hypothesis was not rejected (the point estimate was -0.005 and the *p* value was 0.569). I therefore removed distance to Berlin and estimated the model without this variable using both distance to Wittenberg and the share of Protestants in 1816 as instruments for the share of Protestants in 1871. Table 3 shows the results.

The first-stage *F* statistic for the IV estimate of the effect of Protestantism in equation (3.1) is extremely large, reflecting the high correlation between the share of Protestants in 1816 and the share of Protestants in 1871 (0.981). But distance to Wittenberg is not a redundant instrument: the null hypothesis that it is was rejected with a *p* value of 0.007, so the asymptotic efficiency of the estimation is improved by using distance to Wittenberg as an instrument. Thus, if the regression model does not omit relevant variables that are correlated with distance to Wittenberg, and BW's arguments that distance to Wittenberg is in principle a valid instrument for Protestantism are accepted, the *J* test of overidentifying restrictions is a test of the validity of the share of Protestants in 1816 as an instrument for the share of Protestants in 1871. The *p* value of this test reported in Table 3 (0.343) suggests that it is justifiable to regard the share of Protestants in 1816 as a valid instrument for the share of Protestants in 1871.

[19] Becker et al. (2011), 105 n.12, and online Appendix A1 (https://assets.aeaweb.org/assets/production/articles-attachments/aej/mac/app/2010-0021_app.pdf).

Table 3: The relationship between the literacy rate and Protestantism, Prussia 1871

	Dependent variable: literacy rate	
Regressors	IV	OLS
	(3.1)	(3.2)
Share of Protestants	0.091***	0.103***
	(0.017)	(0.017)
Brandenburg	10.553***	10.667***
	(3.022)	(3.022)
Pomerania	11.141***	10.783***
	(3.055)	(2.970)
Posen	-7.367**	-7.175*
	(3.663)	(3.684)
Silesia	16.299***	16.336***
	(2.913)	(2.909)
Saxony	13.879***	13.785***
	(3.840)	(3.792)
Westphalia	12.186**	12.803***
	(4.914)	(4.993)
Rhineland	11.063**	11.774**
	(5.272)	(5.374)
Year in which annexed by Prussia	0.364	0.378
	(0.297)	(0.290)
Distance to London	-0.020***	-0.020***
	(0.007)	(0.007)
Distance to nearest provincial capital	-0.064***	-0.063***
	(0.012)	(0.012)
Constant	128.882***	122.60***
	(27.789)	(27.674)
Adjusted R^2	0.735	0.735
First-stage F statistic	1467.58	
J test p value	0.343	
C test p value	0.097	
Standard 95 per cent confidence interval for share of Protestants	[0.058, 0.124]	[0.069, 0.137]
Weak-instrument-robust 95 per cent confidence interval for share of Protestants	[0.063, 0.129]	

Notes: Number of observations is 334. Figures in parentheses are heteroscedasticity-robust standard errors. ** and *** denote significance at the 0.05 and 0.01 levels respectively. The results reported for year of annexation and province are the marginal effects for each variable, calculated in the same way as those in Table 1. The omitted province dummy variable is that for the province of Prussia.

The causal effect of Protestantism on literacy in equation (3.1) is precisely estimated, as shown by both the standard and weak-instrument-robust 95 per cent confidence intervals. The latter confidence interval was obtained using the conditional likelihood ratio procedure, as recommended by Andrews et al. (2006) for situations in which the regression model is overidentified, and was computed by the Stata *weakiv* command of Finlay et al. (2013). However, the size of this effect is very small: the

point estimate of the coefficient of the share of Protestants in equation (3.1) corresponds to an elasticity of only 0.064 at sample mean values. Thus there is evidence in favour of the first component of the BW thesis, but the positive causal effect of Protestantism on literacy is a very small one. The estimate in equation (3.1) is approximately half the size of BW's estimate in equation (2) of their Table III. This very small causal effect means that the first component of the BW claim that Protestantism had a positive effect on economic outcomes is a weak one. There is evidence of a positive causal effect of Protestantism on literacy, but it seems to have been extremely small.

Table 3 also reports the *p* value of the *C* statistic which tests the null hypothesis that the education variable being treated as endogenous in IV estimation is actually an exogenous regressor.[20] The *C* statistic amounts to a test of whether there is a statistically significant difference between the IV and OLS estimates of the coefficient of the education variable, and thus whether IV estimation is required. The *p* value of 0.097 does not conclusively fail to reject the null hypothesis. However, the OLS estimates in equation (3.2) are very similar to the IV estimates in equation (3.1), which suggests that IV estimation is not required and the education variable can be treated as exogenous. The point estimate of the coefficient of the share of Protestants in equation (3.2) corresponds to an elasticity at sample mean values of 0.072.

The estimates of the effect of Protestantism on literacy in Table 3 are smaller than those reported by BW, and the reason for this is that the BW estimate is inconsistent. Distance to Wittenberg is not a valid instrument in the BW regression model because it is correlated with omitted variables. There is clear evidence that the province dummy variables and their interactions with the year of annexation were determinants of the literacy rate and hence should not have been omitted from the BW regression model. The two null hypotheses that, in the regression model estimated for literacy, the coefficients of the seven province dummies were all zero and the coefficients of the seven interaction terms were all zero were both rejected. It is also clear from Table 3 that distance to London and distance to the nearest provincial

[20] Hayashi (2000), pp. 218-21; Baum et al. (2003).

capital influenced the literacy rate and should not have been omitted from the BW regression model.

7. Protestantism, literacy and economic outcomes

I turn now to BW's claim that Protestantism had a positive effect on county economic outcomes due largely to the higher literacy of Protestants. To assess this claim, I estimated regression models of the determinants of the three different measures of county economic outcomes used by BW in which the share of Protestants and the literacy rate in 1871 are both included as regressors. The coefficients of these two variables were estimated by IV, using as instruments the share of Protestants in 1816, education in 1816, and (in two of the three models) distance to Wittenberg. Becker et al. (2011) used education in 1816, measured by enrolment in elementary and middle schools as a share of the population aged from six to 14, as an IV for education at later dates in their analysis of the role of education in Prussian industrialisation. They argue that differences in education levels among Prussian counties in 1816 reflected exogenous historical idiosyncracies and therefore had no direct effect on subsequent industrialisation. Edwards (2017) shows that, in order for education in 1816 to be a valid instrument for this purpose, it is necessary for the regression model of industrialisation to include measures of regional effects and the distances to Berlin and London. These variables should, of course, also be included in regression models of county economic outcomes if distance to Wittenberg is to be a valid IV.

I follow Becker et al. (2011) in using several measures of the pre-industrial characteristics of the 334 Prussian counties that existed in 1849 to allow for possible effects of such characteristics on economic outcomes in the later nineteenth century. I include the following variables as regressors for this purpose. The share of the population living in cities in 1816 and the number of looms per capita in 1819 are included as indicators of pre-industrial development. The number of sheep in 1816 is used as a proxy for the availability of wool for the textile industry. The share of farm labourers in the population in 1819 is included as an indicator of whether a county was less likely to industrialise because of its more agricultural orientation. Various measures of pre-industrial public infrastructure which might have influenced

subsequent economic development are also included as regressors: the number of public buildings per capita in 1821, a dummy variable registering the presence of paved inter-regional roads in 1815, and a measure of the capacity of river transport ships in 1819. I also include the geographical area of each county as a regressor.

As in the regression model for literacy, I do not include BW's demographic control variables as regressors. Economic conditions may influence demographic and migration behaviour, and hence population size, population growth, average household size as well as the shares of the population that were below 10, native-born, and Prussian are all potentially endogenous variables in a regression model of county economic outcomes. These variables are likely to be bad controls.

Thus the general regression model used to analyse the effect of Protestantism on economic outcomes for the sample of 334 counties is as follows:

$$Y_i = \beta_0 + \beta_1 P_i + \beta_2 L_i + \beta_3' PR_i + \beta_4 A_i + \beta_5' PR_i * A_i + \beta_6' D_i + \beta_7' X_i + v_i$$

where Y_i denotes a measure of county economic outcome, X_i is a vector of pre-industrial and other characteristics as described above, v_i is an error term, and other variables are as previously defined. I estimated three versions of this model, using BW's three different measures of economic outcomes – income tax per capita, the log of teacher income, and the share of the labour force in manufacturing and services – as dependent variables. In each case, the share of Protestants in 1871 and the literacy rate in 1871 were instrumented by the share of Protestants in 1816 and education in 1816 only: since distance to Berlin was included as a regressor, distance to Wittenberg was an unusably weak instrument. For each version, I tested the null hypothesis that the coefficient of the distance to Berlin variable was zero. This was not rejected at the 0.05 level when the dependent variables were income tax per capita and log teacher income. For these two dependent variables, I therefore estimated regression models that excluded distance to Berlin as a regressor and in which the share of Protestants in 1816, education in 1816 and distance to Wittenberg were used as instruments. However, the model for the share of the labour force in manufacturing and services included distance to Berlin as a regressor and only the share of Protestants and education in 1816 were used as instruments. The results are shown in Table 4.

Table 4: The effects of Protestantism and literacy on county economic outcomes in Prussia

Regressors	Dependent variable: per capita income tax		Dependent variable: ln(teacher income)		Dependent variable: share of labour force in manufacturing and services	
	IV (4.1)	OLS (4.2)	IV (4.3)	OLS (4.4)	IV (4.5)	OLS (4.6)
Share of Protestants in population 1871	-0.129	-0.136	0.0016***	0.0011***	0.164***	0.086***
	(0.184)	(0.120)	(0.0004)	(0.0002)	(0.059)	(0.032)
Literacy rate 1871	0.580	0.584***	-0.0038	-0.0002	-0.667	0.074
	(1.854)	(0.222)	(0.0032)	(0.0006)	(0.464)	(0.076)
Brandenburg	25.510	25.390	0.184***	0.150***	15.047*	7.902*
	(29.608)	(22.173)	(0.048)	(0.034)	(7.852)	(4.464)
Pomerania	18.127	18.293	0.060	0.027	6.995	-0.439
	(28.906)	(17.007)	(0.042)	(0.026)	(6.742)	(3.095)
Posen	-37.024**	-37.083***	-0.007	0.014	-5.555	-1.238
	(14.791)	(11.293)	(0.034)	(0.020)	(4.073)	(1.904)
Silesia	-4.712	-4.838	0.184***	0.122***	17.973*	5.392*
	(36.654)	(13.350)	(0.058)	(0.023)	(9.507)	(3.219)
Saxony	-15.589	-15.687	0.167***	0.127***	10.664	2.592
	(34.749)	(25.751)	(0.053)	(0.040)	(8.768)	(5.190)
Westphalia	-77.057*	-77.497**	0.214***	0.165***	-5.069	-13.236*
	(44.201)	(37.137)	(0.068)	(0.054)	(10.924)	(7.608)
Rhineland	-82.517*	-82.030**	0.256***	0.214***	-11.168	-17.441*
	(43.885)	(39.637)	(0.069)	(0.059)	(12.035)	(9.577)
Year in which annexed by Prussia	-2.559**	-2.571***	-0.0001	-0.0016	-0.028	-0.312***
	(1.027)	(0.868)	(0.0017)	(0.0019)	(0.293)	(0.117)
Distance to London	-0.142**	-0.142***	-0.0002*	-0.0001	-0.056***	-0.039***
	(0.058)	(0.452)	(0.0001)	(0.0001)	(0.015)	(0.105)
Distance to nearest provincial capital	-0.099	-0.099	-0.0002	0.0000	-0.013	0.026**
	(0.112)	(0.085)	(0.0002)	(0.0001)	(0.025)	(0.012)
Distance to Berlin					0.029*	0.027**
					(0.015)	(0.013)

County area	-28.956***	-29.007***	-0.045***	-0.039***	-2.883	-1.624
	(10.279)	(9.406)	(0.017)	(0.015)	(1.960)	(1.490)
Paved streets 1815	23.566**	23.563***	0.062***	0.055***	4.924**	3.448**
	(9.378)	(8.747)	(0.016)	(0.015)	(2.020)	(1.732)
Public buildings per capita 1821	-1,861.288	-1,852.459	-5.346**	-6.118***	-641.778***	-829.196***
	(1,515.938)	(1,416.516)	(2.370)	(2.236)	(232.479)	(196.167)
Sheep per capita 1816	-9.637	-9.673	-0.024*	-0.027**	-5.109***	-5.695***
	(8.591)	(8.672)	(0.013)	(0.012)	(1.682)	(1.555)
Tonnage of ships per capita 1819	201.305	201.461	0.251***	0.209***	22.962	14.484
	(168.979)	(168.861)	(0.083)	(0.078)	(18.008)	(12.917)
Looms per capita 1819	38.346	39.169	0.613*	0.872***	139.500**	189.879***
	(190.418)	(129.798)	(0.347)	(0.229)	(58.161)	(50.353)
Share of farm labourers in population 1819	1.156*	1.154*	0.0020	0.0018	-0.046	-0.082
	(0.690)	(0.692)	(0.0013)	(0.0013)	(0.159)	(0.144)
Share of population living in cities 1816	0.983***	0.984***	0.0033***	0.0032***	0.205***	0.173***
	(0.361)	(0.317)	(0.0005)	(0.0004)	(0.048)	(0.035)
Constant	310.550	313.526**	7.030***	6.599***	115.772*	21.793
	(299.787)	(144.112)	(0.473)	(0.226)	(68.176)	(26.476)
Observations	321	321	334	334	334	334
Adjusted R^2	0.490	0.490	0.738	0.762	0.512	0.632
SW conditional *F* statistic for share of Protestants	14.68		16.55		32.83	
SW conditional *F* statistic for literacy rate	7.29		9.81		19.69	
Standard 95 per cent confidence interval for share of Protestants	[-0.546, 0.287]	[-0.373, 0.101]	[0.0007, 0.0025]	[0.0007, 0.0015]	[0.030, 0.298]	[0.024, 0.149]
Weak-instrument-robust 95 per cent confidence interval for share of Protestants	[-0.546, 0.343]		[0.0009, 0.0027]		[0.030, 0.369]	
Standard 95 per cent confidence interval for literacy rate	[-3.612, 4.773]	[0.149, 1.020]	[-0.0110, 0.0034]	[-0.0013, 0.0010]	[-1.717, 0.383]	[-0.074, 0.222]
Weak-instrument-robust 95 per cent confidence interval for literacy rate	[-5.289, 5.891]		[-0.0149, 0.0024]		[-2.697, 0.243]	
J test *p* value	0.997		0.479		-	
C test *p* value	0.985		0.036		0.129	

Notes: Figures in parentheses are standard errors clustered at the level of the 280 independent units of observation in 1816. *, ** and *** denote significance at the 0.10, 0.05 and 0.01 levels respectively. The results reported for year of annexation and province are the marginal effects for each variable, calculated in the same way as those in Tables 1 and 3. The omitted province dummy variable is that for the province of Prussia.

The presence of two endogenous variables – the share of Protestants in 1871 and the literacy rate in 1871 – in the regressions reported in Table 4 means that the conventional first-stage *F* statistic cannot be used to test the weakness of the instruments. Table 4 therefore reports the Sanderson-Windmeijer (2016) conditional first-stage *F* statistics for the share of Protestants and the literacy rate, which suggest that weak instrument problems have to be taken seriously, especially for the literacy rate in equations (4.1) and (4.3).

Table 4 reports 95 per cent weak-instrument-robust confidence intervals for the estimated effects of the share of Protestants and the literacy rate, calculated using the Stata *weakiv* command of Finlay et al. (2013). For equations (4.1) and (4.3), these confidence intervals were obtained from a grid search using Kleibergen's (2005) GMM extension of the conditional likelihood ratio test to find the set of values of the coefficients of the two endogenous variables such that the null hypothesis that both these coefficients were zero could not be rejected.[21] The confidence interval for equation (4.5) was also obtained from a grid search, but one using the Anderson-Rubin (1949) test, since this is the only test that can be employed in just-identified models. The confidence intervals reported in Table 4 are the ranges of values of the relevant coefficients that were included in the sets obtained by grid search.

For all three equations estimated by IV in Table 4, the difference between the weak-instrument-robust and standard confidence intervals for the literacy rate is much larger than the corresponding difference for the share of Protestants. The estimated effect of the literacy rate is poorly determined in all three of equations (4.1), (4.3) and (4.5), particularly in equation (4.1). This is not because (4.1) was estimated on a smaller sample than the other two equations: the relatively greater precision of the estimated effect of the literacy rate in (4.3) and (4.5) persists if these equations are estimated on the sample of 321 for which per capita income tax data are available. Equations (4.1), (4.3) and (4.5) were all estimated by two-stage least squares, which is an IV estimator that can perform poorly when instruments are weak. However, when I used an IV estimator that has good properties in such circumstances - Fuller's modified limited information maximum likelihood procedure with parameter one

[21] I thank Mark Schaffer for clarifying this point.

(Fuller 1977) - the results changed so little that two-stage least squares estimates appear to be acceptable in this case.

The fact that the conditional *F* statistics for equation (4.5), which does not use distance to Wittenberg as an IV, are larger than those for equations (4.1) and (4.3), which do, might be thought to suggest that distance to Wittenberg is the source of the weak instrument problems and consequently should not be used to estimate (4.1) and (4.3). However, comparison of the standard and weak-instrument-robust confidence intervals for the estimated effects of the share of Protestants and the literacy rate in equation (4.5) shows that this equation is not free of weak-instrument problems. Furthermore, the null hypothesis that distance to Wittenberg is a redundant instrument is rejected with *p* value 0.025 in equation (4.1) and 0.041 in equation (4.3). Thus there is a strong case for using distance to Wittenberg as an IV to estimate these two equations in order to improve the asymptotic efficiency of the estimation. Doing so also means that the *J* test of overidentifying restrictions can be used to test whether, conditional on accepting BW's argument that distance to Wittenberg is a valid instrument for the share of Protestants in 1871 and Becker et al.'s argument that education in 1816 is a valid instrument for the literacy rate in 1871, the share of Protestants in 1816 is also a valid instrument. The *p* values for the *J* test reported for equations (4.1) and (4.3) in Table 4 show that, provided the other two IVs are valid, the share of Protestants in 1816 is a valid instrument.

Table 4 shows a clear difference between the results when per capita income tax is the measure of county economic outcomes and when the other two measures are used. The point estimate of the effect of Protestantism in equation (4.1) is negative while that for literacy is positive, but neither effect is well-determined, with the weak-instrument-robust confidence interval for literacy being extremely large. However, the *p* value of the *C* statistic for (4.1) is so large as to suggest that there is no need for IV estimation in this case. Of course, when the IV estimates are poorly-determined, the *C* test may have low power, but the OLS point estimates in equation (4.2) are so similar to the IV ones in (4.1) that there is very little danger of making a type II error in concluding that IV estimation of the regression model with per capita income tax as dependent variable is not required. The OLS point estimate of the effect of Protestantism on this measure of county economic outcomes corresponds to an

elasticity of -0.043 at sample mean values: it is therefore of essentially no economic significance as well as being not statistically significantly different from zero. In contrast, the OLS point estimate of the effect of literacy on per capita income tax is statistically significant and of modest economic significance, corresponding to an elasticity at sample mean values of 0.260.

However, the estimated effects of Protestantism and literacy change when county economic outcomes are measured by either the log of teacher income or the share of the labour force in manufacturing and services. For the former measure, the *C* test rejects the null hypothesis of no difference between the IV and OLS estimates, so IV estimation is required. The 95 per cent weak-instrument-robust confidence interval for the estimated effect of Protestantism shows that this is positive and statistically significantly different from zero, though its economic significance is small: the point estimate in (4.3) corresponds to an elasticity of 0.094 at sample mean values. The point estimate of the effect of literacy corresponds to an elasticity of -0.325 at sample mean values, but is not statistically significant.

The *p* value of the *C* statistic for equation (4.5), in which the dependent variable is the share of the labour force in manufacturing and services, is 0.129, so the null hypothesis of no difference between IV and OLS estimates cannot be rejected at conventional levels. But in this case one must take seriously concerns about the low power of the *C* test when IV estimates are not well-determined. The IV and OLS point estimates of the effects of Protestantism and literacy differ substantially, as do many of the other coefficient estimates in equations (4.5) and (4.6). The possibility of making a type II error by concluding that IV estimation is not required is sufficiently great that I focus on the results of equation (4.5). The 95 per cent weak-instrument-robust confidence interval for the estimated effect of Protestantism in this equation shows that it is positive and statistically significantly different from zero. The point estimate corresponds to an elasticity at sample means of 0.299, so the economic significance of Protestantism, though modest, is non-trivial. The point estimate of the effect of literacy is much larger (in absolute value), corresponding to an elasticity at sample mean values of -1.723, but it is so poorly determined as to be statistically insignificant.

Is it plausible that the effect of literacy on log teacher income and share of labour force in manufacturing and services might have been negative? A possible explanation of the negative causal effect of literacy on the latter measure is that greater literacy in 1871 reflected increased education in the decades before that date, which reduced the supply of child labour to factories, thus increasing the cost of labour and lowering the profitability of industrial activity, so that the share of the labour force in manufacturing was lower in 1871. Some support for this explanation comes from the debates preceding the enactment of the Prussian child labour law in 1839: many opponents of this new legislation were concerned that removing children from their jobs in order to send them to school would be damaging to industry (Anderson 2013). A negative effect of literacy on the share of manufacturing occupations is therefore consistent with contemporary evidence on the Prussian economy. Whether this explanation can also apply to teacher income is not clear. On the one hand, a negative effect of literacy on industrialisation might lower the general level of incomes in a county, including teacher income. On the other hand, increased education might be expected to raise the relative income of teachers due to greater demand for them. In any event, a negative effect of literacy on this measure of county economic outcomes is not impossible.

In all the regression models reported in Table 4 there is strong evidence that regional effects had an important influence on county economic outcomes. The null hypothesis that the coefficients of the province dummy variables, the date of annexation by Prussia, and the terms that interacted these variables were all zero was rejected with p values of 0.000 in all equations except (4.5), for which it was rejected at the 0.046 level. The fact that these variables cannot be excluded from the regressions in Table 4 reinforces the point made earlier: the estimates of the effect of Protestantism from BW's regression models of county economic outcomes would be inconsistent even if Protestantism had no causal effect on literacy, because distance to Wittenberg is correlated with variables that influence economic outcomes but were omitted from BW's regressions.

There is also strong evidence that both county area and the pre-industrial characteristics of counties that were included in the regression models in Table 4 had an important influence on county economic outcomes. The null hypothesis that the

coefficients of these variables were all zero was rejected with *p* value 0.000 in all six equations in Table 4. There is no particular reason to suppose that the omission of these variables from BW's regression models of economic outcomes creates additional problems for the consistency of their IV estimates, but including them increases the likelihood that the instruments used satisfy the exclusion restriction.

The conclusions to be drawn from Table 4 about the effects of Protestantism and literacy on county economic outcomes differ according to which measure of these outcomes is used. However, the results in Table 4 reject the BW view that Protestantism had a positive effect on economic outcomes which was largely due to the causal effect of Protestantism on literacy. It might be thought that the results using BW's preferred measure of economic outcomes – per capita income tax – support the BW view, because equation (4.2) shows that Protestantism had an effect which was poorly determined and essentially zero, while literacy had a clear positive effect, though one of a modest size. But the BW view requires it to be the case that Protestantism had a positive effect if no account is taken of literacy. If literacy is omitted from equations (4.1) and (4.2), the resulting coefficient estimate for the share of Protestants is -0.083 in the former and -0.082 in the latter, so that the omission of literacy fails to make the estimated effect of Protestantism positive as the BW view would require.[22]

The estimated causal effects of Protestantism and literacy when log teacher income and share of labour force in manufacturing and services are used to measure economic outcomes are broadly similar. They also show that the BW view is untenable. The causal effect of Protestantism is clearly positive, though very small in equation (4.3) while larger, but still modest, in equation (4.5). The causal effect of literacy is not well-determined, but the point estimates are negative in both equations (4.3) and (4.5), modestly so in the former and substantially so in the latter. Protestantism had an unambiguous positive effect on both these economic outcome measures, and literacy had an ambiguous effect which was, if anything, negative. Such effects cannot be reconciled with the BW view, which requires there to be an

[22] The *p* value of the *C* test in this case is 0.981 and shows that, even with literacy omitted, there is no need for IV estimation. The 95 per cent confidence interval for the OLS estimate of the share of Protestants when literacy is omitted is [-0.321, 0.157].

unambiguous positive effect of literacy that is substantially larger than any positive effect of Protestantism.

8. Conclusion

This paper has shown that late nineteenth-century Prussia provides no evidence to support the argument that Protestantism had a positive effect on economic prosperity which can largely be explained by the higher human capital of Protestants. BW's claim that this is the case is rendered untenable by fundamental problems with their empirical strategy. It is simply not possible to obtain consistent estimates of the respective causal effects of Protestantism and literacy in the absence of an IV for literacy. This paper obtained such an IV by combining the BW dataset with another dataset, used by Becker et al. to study education and industrialisation in Prussia. The combined dataset was used to estimate the causal effects of both Protestantism and literacy for a subsample of the Prussian counties considered by BW. In addition to providing IV estimates of the effects of literacy as well as Protestantism, this paper also addressed another problem with the BW empirical strategy: the absence of any measures of regional effects on Prussian economic development in their regression models. The omission of regional effects makes distance to Wittenberg (BW's instrument for Protestantism) an invalid IV.

The results of the empirical analysis in this paper provide evidence that Protestantism had a positive causal effect on literacy in nineteenth-century Prussia. However, the size of this effect is so very small as to make it unlikely, though not impossible, that any positive effect of Protestantism on economic outcomes can be mainly due to the greater literacy of Protestants.

The empirical analysis in this paper does not provide clear evidence that Protestantism had a positive causal effect on county economic outcomes in nineteenth-century Prussia. The results vary according to the economic outcome measure used. There is no evidence of a positive effect of Protestantism on income tax per capita (BW's preferred measure). The effect of Protestantism on teacher income is positive, but small, while the effect on the share of manufacturing and services in the labour force is also positive and large enough to be of some economic

significance. This mixed evidence for nineteenth-century Prussia leaves open the possibility that there is a grain of truth in the Weber thesis in its general form of some causal connection existing between Protestantism and economic prosperity.

However, the empirical analysis in this paper unambiguously rejects the BW view that Protestantism had a positive effect on economic prosperity that was largely due to the greater literacy of Protestants. Even putting to one side the very small size of the causal effect of Protestantism on literacy, the estimated effects of these two variables on county economic outcomes cannot be reconciled with the BW view. Literacy had a positive effect on income tax per capita, but the effect of Protestantism was negative (though poorly determined) irrespective of whether literacy was or was not included as a regressor. For the other two measures of economic outcomes, the effect of Protestantism was positive while that of literacy was negative, though poorly determined. There may be a grain of truth in the Weber thesis, but BW's interpretation of it cannot be sustained.

References

Acemoglu, D., D. Cantoni, S.H. Johnson and J.A. Robinson (2011). 'The Consequences of Radical Reform: The French Revolution', American Economic Review, 101, 3286-3307.

Anderson, E. (2013). 'Ideas in Action: The Politics of Prussian Child Labor Reform, 1817-1839', Theory and Society, 42, 81-119.

Anderson, T.W. and H. Rubin (1949). 'Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations', Annals of Mathematical Statistics, 20, 46-63.

Andrews, D.W.K., M.J. Moreira and J.H. Stock (2006). 'Optimal Two-sided Invariant Similar Tests for Instrumental Variables Regression', Econometrica, 74, 715-52.

Angrist, J.D. and J-S. Pischke (2009). Mostly Harmless Econometrics. Princeton: Princeton University Press.

Baum, C.F., M.E. Schaffer and S. Stillman (2003). 'Instrumental variables and GMM: estimation and testing', The Stata Journal, 3, 1-31.

Becker, S.O. and L. Woessmann (2009). 'Was Weber Wrong? A Human Capital Theory of Protestant Economic History', Quarterly Journal of Economics, 124, 531-596.

Becker, S.O., E. Hornung and L. Woessmann (2011). 'Education and Catch-Up in the Industrial Revolution', American Economic Journal: Macroeconomics, 3, 92–126.

Becker, S.O., F. Cinnerella, E. Hornung and L. Woessmann (2014). 'iPEHD - The ifo Prussian Economic History Database', Historical Methods: A Journal of Quantitative and Interdisciplinary History, 47, 57-66.

Cameron, A.C., J.B. Gelbach and D.L. Miller (2008). 'Bootstrap-based improvements for inference with clustered errors', Review of Economics and Statistics, 90, 414-27.

Cantoni, D. (2012). 'Adopting a New Religion: The Case of Protestantism in 16th Century Germany', Economic Journal, 122, 502-31.

Dufour, J-M. (2003). 'Identification, weak instruments, and statistical inference in econometrics', Canadian Journal of Economics, 36, 767-808.

Edwards, J.S.S. (2017). 'A replication of "Education and Catch-Up in the Industrial Revolution" (American Economic Journal: Macroeconomics, 2011)', Economics Discussion Papers, No 2017-30, Kiel Institute for the World Economy, http://www.economics-ejournal.org/economics/discussionpapers/2017-30.

Finlay, K., L.M. Magnusson, and M.E. Schaffer (2013). weakiv: Weak-instrument-robust tests and confidence intervals for instrumental-variable (IV) estimation of linear, probit and tobit models. http://ideas.repec.org/c/boc/bocode/s457684.html

Fuller, W.A. (1977), 'Some Properties of a Modification of the Limited Information Estimator', Econometrica, 45, 939-53.

Hayashi, F. (2000). Econometrics. Princeton: Princeton University Press.

Kleibergen, F. (2005). 'Testing Parameters in GMM Without Assuming That They Are Identified', Econometrica, 73, 1103-23.

Ogilvie, S.C. (1996a). 'Proto-Industrialization in Germany', in S.C. Ogilvie and M. Cerman (eds.), European Proto-Industrialization, Cambridge: Cambridge University Press.

Ogilvie, S.C. (1996b). 'The Beginnings of Industrialization', in S.C. Ogilvie (ed.), Germany: a New Social and Economic History, Vol. II: 1630-1800, London, Edward Arnold.

Sanderson, E ., and F. Windmeijer (2016). 'A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables', Journal of Econometrics, 190, 212-221.

Tipton, F.B. (1976). Regional Variations in the Economic Development of Germany During the Nineteenth Century. Middletown: Wesleyan University Press.